J. Clin. Chem. Clin. Biochem.
Vol. 27, 1989, pp. 613–629

© 1989 Walter de Gruyter & Co.
Berlin · New York

# Performance Profiles: New Tools for Characterization and Comparison of Clinical Chemical Results

By *H. Keller*

*Institut für Klinische Chemie und Hämatologie des Kantons St. Gallen* and

*H. Passing*

*Hoechst AG, Abteilung Informatik und Kommunikation/Software, Frankfurt/Main*

*Dedicated to Professor Konrad Spang, former director of the medical clinic of the Katharinen-Hospital Stuttgart on the occasion of his 80th birthday*

(Received January 13/June 26, 1989)

**Summary:** It is current practice to record the performance of immunoassays by means of precision profiles (according to *Ekins*), in which the variation coefficient (relative standard deviation) is plotted against the analyte concentration. On the other hand, precision profiles are only occasionally used for evaluating the performance of conventional clinical-chemical methods.

It is relatively uncommon to find bias plotted against analyte concentration, presumably because this type of analysis requires reference specimens, whose true analyte concentrations are known or guaranteed by reference methods. If the relative systematic variations are plotted against the true analyte concentrations, and a confidence interval is added to the resulting regression curve, the result is a "bias profile"; if tolerance limits are added, the result is a "deviation profile".

The present work describes the preparation of specimens, which can be used to provide experimental data for the three performance profiles. A computer program is used to construct the precision profile, bias profile and/or deviation profile. The mathematical-statistical basis of the program is described in detail.

For evaluation of the statistical procedure, three analytes and six methods were used: determination of sodium activity/concentration with an ion sensitive electrode and by flame photometry; determination of creatinine by a manual enzymic and a mechanized *Jaffé* method; determination of thyrotropin by radioimmunoassay and by luminescence immunoassay.

Different purposes are served by bias and deviation profiles. Thus, bias profiles can be used to compare the bias of two methods, whereas a deviation profile can be used to define the analytical range of a method. If the acceptable limits of deviation are added to the deviation profile, then the useful analytical range of the method is immediately apparent.

## Introduction

In a study of the reliability criteria of chemical analytical methods, *Wilson* (2) refers to a 1963 monograph by *Nalimov* (1) and states: "A further point of importance is the possible dependence of s on the concentration of determinand". In clinical chemistry, the dependence of analytical precision and accuracy on the concentration of the analyte in the matrix was

apparently first described and analysed graphically in 1967 (3). In 1971, *Broughton* et al. (4) evaluated a number of autoanalyser methods and showed, in practically every case, that an increase in the concentration of an analyte was accompanied by an increase in the standard deviation and a decrease in the variation coefficient. In 1974, *Aronson* et al. (5) reported similar studies with six "automated" analytical systems. In all the tested methods, the standard deviation increased with the concentration of analyte, whereas the variation coefficient at first decreased, then increased again at higher analyte concentrations.

In the same year, *Rodbard* (6) investigated the dose-response curves of various radioimmunoassays, and reported that the relative standard error depended on the concentration. If the standard deviations or variation coefficients are plotted against the analyte concentrations in an x/y diagram, and the points are fitted to a curve, the resulting plot is the "precision profile" described by *Ekins* (7). It is possible to construct intra-assay/intra-sample, intra-assay/inter-sample, inter-assay/intra-sample and other forms of precision profile. *Ekins* (8) recommended the use of these curves for optimizing the reagents, the experimental procedures and the instrumental settings. In recent years, many authors have used precision profiles to describe the performance of immunoassays (9–12).

Hitherto, the plot of $\bar{x}_i/s_i$ (or $\bar{x}_i/CV_i$), has been drawn "intuitively" (i.e. manually). Recently, *Sadler* et al. (13) formulated a model, which permits an approximate estimation or graphic representation of variance:

$$\sigma^2 (H) = (\beta_1 + \beta_2 H)^I \qquad \text{(Eq. 1)}$$

where $\sigma^2(H)$ represents variance, H represents the concentration, and $\beta_1$, $\beta_2$ and I are parameters.

It has been shown that a good approximation results if I is assigned a value of 3.

In general, the following can be stated: if H = 0, the variance corresponds to the background "noise" of the analytical system, i.e. imprecision of the blank value. At measurable analyte concentrations, random errors (imprecision) are accordingly always the same or larger than the scatter around the blank value.

Referring to the deviations from accuracy (bias), *Mandel* (14) states: "depends on two factors: the concentration of the analyte in the sample, and the matrix in which this analyte finds itself in the sample ... Consequently, the accuracy of an analytical method should never be determined on a single sample, but rather on a set of samples covering the entire range of concentrations of the analyte and representing as many as possible of the matrix configurations that may typically be expected in the situation under study".

The analytical calibration function, starting from a "single component model" was defined by *Currie* (15) as:

$$\hat{x} = (y - B)/A \qquad \text{(Eq. 2)}$$

where B = blank, A = calibrator factor, $\hat{x}$ = estimated concentration, y = instrument response.

The "proportional bias" causes A to change, whereas the "constant bias" causes B to change. In general, however, the following applies: "the inaccuracy is understood to ... consist of two components, the bias and the imprecision ... any particular combination to replace bias and STD with a single index ... is not recommended; it must be viewed as a two-component vector". The concept of "total errors" (see, e.g. l.c. (26)) should be abandoned.

If, for the purposes of determining bias, the analyte concentration in specimen i is determined n times, then the average value, $\bar{x}_i$, is subject to analytical imprecision. The difference $(\bar{x}_i - \mu_i)$ must therefore be supplemented with a confidence interval, which can be estimated (14):

$$(\bar{x} - \mu) - t_\alpha \frac{s}{\sqrt{n}} < \text{bias} < (\bar{x} - \mu) + t_\alpha \frac{s}{\sqrt{n}} \qquad \text{(Eq. 3)}$$

where $t_\alpha$ is the critical value of the t-distribution (*Student*), for probability 1-α, and n-1 degrees of freedom. $\bar{x}_i$ represents the estimator of the expected value (in the method under investigation), whereas $\mu_i$ represents the true value (or assigned value, or reference value, etc.).

If bias is determined for several analyte concentrations over the total relevant range, a "bias profile" analogous to the "imprecision profile" can be constructed.

If, however, the confidence limits (Eq. 3) are changed to tolerance limits,

$$(\bar{x} - \mu) - t_\alpha s \sqrt{1 + \frac{1}{n}} < \text{deviation} < (\bar{x} - \mu) + t_\alpha s \sqrt{1 + \frac{1}{n}} \qquad \text{(Eq. 4)}$$

and the bias, together with these tolerance limits, is plotted over the relevant concentration range, then a different profile is obtained, for which we propose the name "deviation profile".

The limits of this deviation profile run parallel to those of the bias profile. The limits of the deviation profile are wider than those of the bias profile, because the bias profile represents only the systematic deviation of the method, whereas the deviation profile includes both the systematic and random deviations. This is explained more fully in the appendix.

The profiles of bias and deviation serve different purposes. For example, bias profiles can be used to compare the bias of two methods, whereas a deviation profile can be used to define the useful analytical range of a method. For the latter purpose, the acceptable limits of deviation are drawn on the graph. The range of analyte concentrations lying within these deviation limits (i. e. the useful analytical range of the method) is then immediately apparent. This is also described more fully in the appendix.

Figure 1 shows model graphical representations of these concepts, based on the data in table 1. In figure 1a, standard deviation is plotted against the average value; it is assumed that each average value was calculated from 10 separate measurements. The points are joined by a (slightly convex) curve, corresponding to Eq. 1. If this representation is based on relative values, i. e. the ordinate becomes the percentage variation coefficient, a different plot is obtained (fig. 1b). The bias of these model results is shown by figure 1c, where the points are joined by a straight line corresponding to the regression function of table 1. The accompanying standard deviations are determined according to Eq. 3[1]). The resulting "bias profile" is hatched in figure 1c, and it can be seen that this becomes slightly wider as $\mu$ increases. If the relevant standard deviations are determined according to Eq. 4[2]), the tolerance limits (dashed lines) are obtained. This "deviation profile" is about 3.3 times wider than the bias profile.

The relative presentation of the same material is shown in figure 1d. It contains all the information of the three previous graphs, and shows at a glance the limits of performance of the analytical method. This model can be translated into practice as follows:

1. Several (at least 3) specimens must be available, whose analyte concentrations ($\mu$) are known exactly, e. g. control specimens are prepared by weighing the pure analyte, or their concentrations are determined with a definitive or reference method. The analyte concentrations of these specimens must cover either the whole of the analytically relevant concentration range, or that part of the range intended for current use.

Tab. 1. Model population of results. The average values ($\bar{x}$) were calculated from the regression function, and the standard deviation (s) was determined according to l. c. (13).
$\mu$ = true value; CV = coefficient of variation; CB = coefficient of bias

| $\mu$ (Arb. units) | $\bar{x}$ (Arb. units) | s (Arb. units) | CV (%) | CB (%) |
|---|---|---|---|---|
| 10 | 9.5 | 1.043 | 10.98 | −5.0 |
| 30 | 30.0 | 1.138 | 3.79 | 0 |
| 50 | 50.5 | 1.236 | 2.45 | 1.0 |
| 70 | 71.0 | 1.336 | 1.88 | 1.43 |
| 90 | 91.5 | 1.439 | 1.57 | 1.66 |

Regression function: $\bar{x} = 1.025\,\mu - 0.75$; $\beta_1 = 1.0$; $\beta_2 = 0.003$

2. Using the method under test, the chosen analyte is determined several times (e. g. 10 times) in each specimen. $\bar{x}$ and s are calculated from the resulting data, thereby confirming (e. g. by means of the *Kolmogorov-Smirnov* test) that the data in fact show a normal distribution.

3. Using the average values and the standard deviations or variation coefficients, the parameters $\beta_1$ and $\beta_2$ are estimated with the aid of Eq. 1, and a precision profile is constructed.

4. Using Eq. 3 and the average and assigned values, the bias and the accompanying interval are calculated for each concentration, and the "bias profile" is plotted.

5. Using Eq. 4 and the average and assigned values, the bias and the accompanying tolerance interval are calculated for each concentration, and the "deviation profile" is plotted. The acceptable limits of deviation are added to this graph, so that the useful analytical range of the method can be read off.

6. Using other specimens with different matrices, these experimental procedures are repeated, preferably by several different laboratories.

Using this procedure, the following characteristics of the method are determined:

a) the quantitative relationship between imprecision and the assigned (true) value;

b) the quantitative relationship between bias and the assigned (true) concentration; and

c) the upper and lower limits of the useful analytical range, which depend on the acceptable limits of deviation.

[1]) $s \times t_\alpha/\sqrt{n} = s \times 0.715$;

[2]) $s \times t_\alpha \sqrt{1 + 1/n} = s \times 2.373$;
$t_\alpha$ (two-sided, for $n - 1 = 9$) = 2.2622
(see, e. g. Scientific Tables Geigy, 8th edition, Vol. 3. Significance limits of *Student*'s distribution).

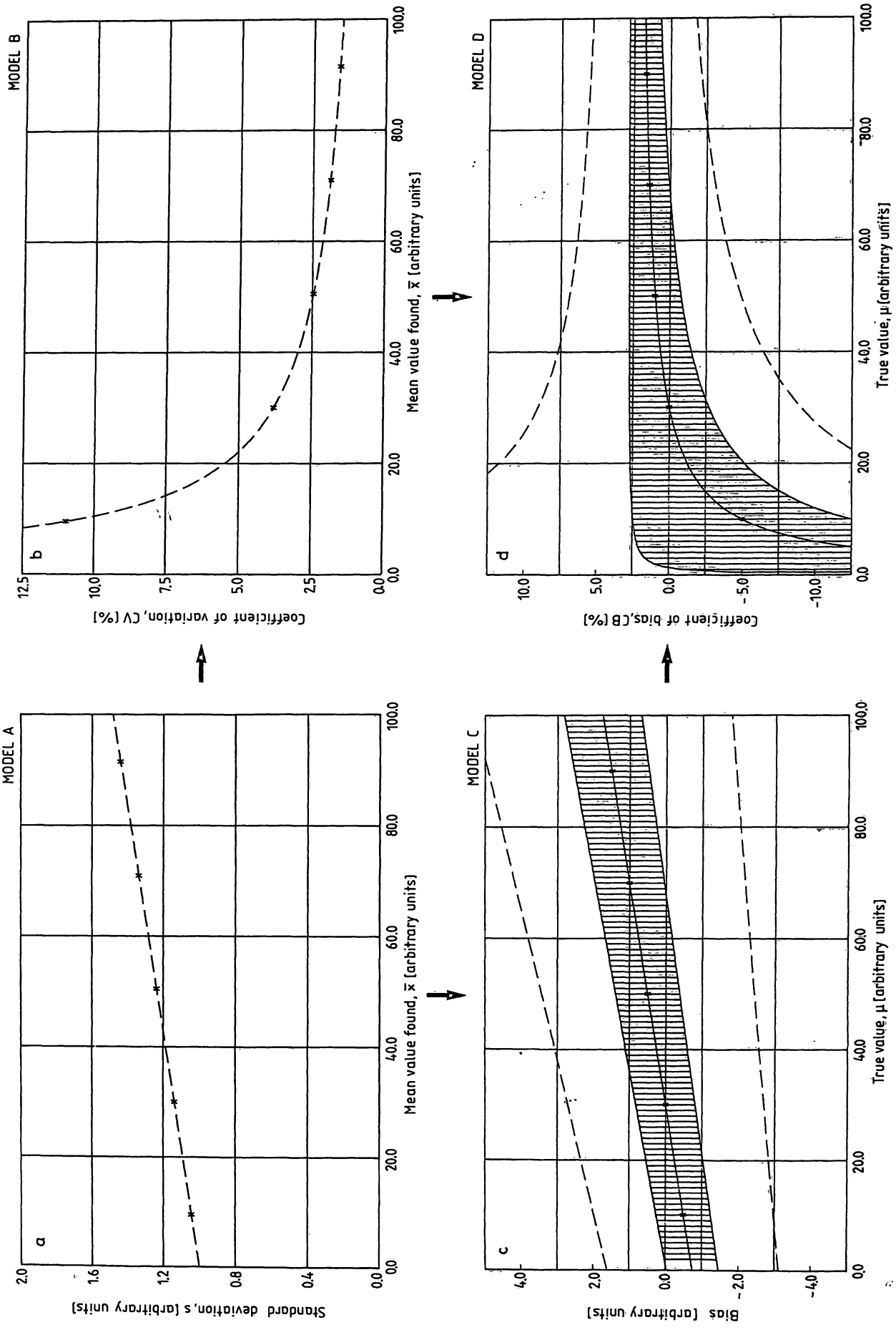


Fig. 1. Graphical representations of the model results in table 1.
a) Model A, mean value found, $\bar{x}$ vs standard deviation, s;
b) Model B, mean value found, $\bar{x}$ vs relative standard deviation (coefficient of variation), CV;
c) Model C, true value, μ vs bias with intervals according to equations 3 and 4;
d) Model D, true value, μ vs relative bias (coefficient of bias), CB.
The bias profile (from Eq. 3) is hatched, and the tolerance limits (from Eq. 4) are shown as dashed lines.

We have tested this concept on three different analytes, using different analytical procedures.

## 2. Materials and Methods

The analytical apparatus and commercial test kits employed in this work were used strictly in accordance with the instructions of the manufacturers. Calibrated apparatus was used for all volumetric work.

### 2.1 Sodium determination in aqueous NaCl solutions

Two analytical systems were used for the determination of sodium concentration or activity:

#### 2.1.1 *Automatic flame photometer, Eppendorf AFM 5051*

The apparatus automatically determines the concentrations of sodium, potassium and calcium in serum, and sodium and potassium in urine. Undiluted specimens are introduced into the analyser in closed vessels. Dilution, mixing and measurement are performed by the analyser, and the results are printed out immediately after the analysis. An acetylene flame is used for excitation. The intensity of light emission is measured after selection of the appropriate wavelengths by interference filters.

The diluent contains a constant concentration of lithium, so that the analyser determines the ratio of the two measurement signals, e. g. sodium/lithium, and compensates for any changes of concentration that may have occurred during atomization or introduction of the gases. The instrument dilution system mixes 1 part of sample with 50 parts of diluent. Standard solution No. 0030 310.004 from the manufacturer was used for the calibration.

#### 2.1.2 *Ion-sensitive electrode unit in the Hitachi 717 automatic analyser, Boehringer Mannheim*

Ion-sensitive electrodes (ISE) are constructed as flow-through electrodes with fluid membranes on a PVC base. They contain a crown ether for the determination of sodium, valinomycin for the determination of potassium, and a quaternary ammonium compound for the determination of chloride. The analyser pipettes 20 µl of sample into a dilution vessel, where it is mixed with 600 µl of diluent (No. 820 636). Finally, the diluted sample is sucked into the ISE compartment until the electrodes are covered by the solution. The activities of sodium, potassium and chloride are always measured simultaneously.

The manufacturer quotes the following measurement ranges:

sodium 80 – 180 mmol/l,
potassium 1.5 – 10 mmol/l,
chloride 60 – 120 mmol/l.

Before and after each sample measurement, the ISE system is recalibrated with an internal standard solution (No. 820 644). Once daily, before operations begin, calibration is performed with two standard solutions (low, No. 646 911; high, No. 646 938), and a compensator (No. 917 656).

#### 2.1.3 *Preparation of test material*

A series of sodium chloride dilutions was prepared as follows. The contents of an ampoule of sodium-standard solution Titrisol® (Merck Darmstadt, No. 9927), containing 1.000 g sodium as sodium chloride, was washed into a 250 ml volumetric flask with double distilled water and adjusted to the calibration mark. The resulting solution contained 173.989 mmol/l $Na^+$. To 10 ml samples of this solution were added 0, 1, 2, 3 ... 10 ml of double distilled water, and the concentration of each dilution was calculated from: 173.989 mmol/l per total volume (ml) $\times 10^{-1} = c_{Na}$ mmol/l.

### 2.2 Determination of creatinine in mixed sera

Two different methods were used for the determination of creatinine:

#### 2.2.1 *Manual determination of the creatinine concentration*

In the test kit "Creatinine PAP" (Boehringer Mannheim, No. 839 434) creatinine is hydrolysed to creatine by the action of creatininase (EC 3.5.2.10). Creatinase (EC 3.5.3.3) then catalyses the conversion of creatine to sarcosine and urea. In a third stage, sarcosine is converted to glycine, formaldehyde and hydrogen peroxide by the action of sarcosine oxidase (EC 1.5.3.1). Finally, in the presence of peroxidase, the hydrogen peroxide reacts with 3,5-dichloro-2-hydroxybenzene sulphonic acid and *p*-aminophenazone (PAP) to produce a red benzoquinonimine, which is measured at 546 nm.

#### 2.2.2 *Creatine determination with the analysis system Hitachi 705*

The method is based on a modification of the *Jaffé* reaction. Creatinine forms a coloured complex with picrate in alkaline solution, and the rate of formation of this complex is measured. The test combination from Boehringer Mannheim (No. 704 130) was used for the assay.

#### 2.2.3 *Preparation of test material*

Sera with very low creatinine concentrations (< 50 µmol/l) were selected from the daily samples received by the clinical laboratory, and combined into a serum pool. A second pool was prepared in parallel, using sera with creatinine concentrations between 1000 and 1500 µmol/l. Using an HPLC procedure as the reference method, the low creatinine pool (serum pool 1) was found to have a concentration of 40.66 µmol/l (C1) and the high creatinine pool (serum pool 2) contained 1169.09 µmol/l (C2)[3]).

Separate 4 ml ($v_1$) samples of serum pool 1 were mixed with 0, 50, 100, 150, 200, 250, 300 and 350 ml ($v_2$) of serum pool 2. The creatinine concentrations of the 8 resulting serum mixtures were calculated from: $[(C_1 \times v_1) + (C_2 \times v_2)]/(v_1 + v_2) = C_3$, where the concentration (C) is in µmol/l, and the volume (v) is in ml.

### 2.3 Thyrotropin determination in mixed sera

Two commercial thyrotropin reagent kits were available for the determination of serum thyrotropin. One of these assays uses $^{125}$I-labelled antibody, and the radioactivity was measured in a gamma counter (MR 252, Kontron AG, ZH). In the second assay, the label is a luminogen, and the luminometer, Clinilumat® LB 9501 (supplier: Dr. Berthold, Wildbad), was used for the luminescence measurements.

#### 2.3.1 *Radioimmunoassay of thyrotropin using two monoclonal antibodies (Henning Berlin)*

In this radioimmunoassay, two monoclonal antibodies recognize two different determinants of the antigenic human thyrotropin. One of the two antibodies is labelled with $^{125}$I, and the

[3]) The authors are very grateful to Dr. *Leinberger*, Boehringer Mannheim, for these determinations.

other is bound to the surface of minute plastic particles. About $3.6 \times 10^9$ particles are added to each assay tube, corresponding to a total surface area of 114 $cm^2$. Free, $^{125}I$-labelled antibody, fixed antibody and test material are incubated together at room temperature; the resulting sandwich complex remains fixed to the microparticles. After incubation, the reaction solution is diluted and the bound fraction is sedimented by centrifugation. The radioactivity of the microparticle pellet is proportional to the thyrotropin concentration. The lower limit of the assay is 0.03 mU/l. The WHO standard 2 IRP 80/558 was used as the reference standard.

2.3.2 *Luminescence immunoassay (Byk-Mallinckrodt)*

In this assay, a monoclonal antibody is bound covalently to the inner wall of the assay tube. The antigenic human thyrotropin of the sample or standard binds to the monoclonal antibody. A second monclonal antibody is then added, which is labelled with a luminogenic molecule. During the incubation, a sandwich is formed, which remains in the assay tube when the excess solution is removed by aspiration. After washing, the drained assay tube is placed in the luminometer; catalyst solution and alkaline peroxide solution are added, and the resulting light emission is measured. Using the emission intensities of the standards, the apparatus performs an internal calibration, and uses this to automatically calculate the thyrotropin concentrations of the test samples.

2.3.3 *Preparation of test material*

Sera with relative high thyrotropin concentrations were selected from the daily samples received by the clinical laboratory, and combined into a serum pool (high thyrotropin pool). Similarly, sera with very low thyrotropin concentrations were combined into a low thyrotropin pool. The respective concentrations of the two pools were determined in an independent reference laboratory[4]), using the RIA-gnost TSH (Behring-Werke) as the reference method. The reference standard was TSH WHO 80/558. The high thyrotropin pool was found to contain 46.46 mU/l, and the low thyrotropin pool contained 0.016 mU/l.

From the two pools, 8 mixtures were prepared according to the following scheme:

| No. | High thyrotropin pool | + | Low thyrotropin pool | Theoretical concentration (mU/l) |
|---|---|---|---|---|
| 10 | 1 | + | 0 | 46.46 |
| 9 | 1 | + | 1 | 23.238 |
| | Previous dilution | | Low thyrotropin pool | |
| 8 | 1 | + | 3 | 5.8215 |
| 7 | 1 | + | 9 | 0.5965 |
| 6 | 1 | + | 1 | 0.3063 |
| 5 | 1 | + | 1 | 0.1611 |
| 4 | 1 | + | 1 | 0.0886 |
| 3 | 1 | + | 1 | 0.0523 |
| 2 | 1 | + | 1 | 0.0341 |
| 1 | 0 | + | 1 | 0.016 |

2.4 Procedure for the evaluation of analytical precision and accuracy

In all the tested methods, the calibrators recommended or supplied by the manufacturer were used. These calibrators meet the requirements of *Büttner* et al. (16) and *Stamm* (17), i.e. that calibrators should be independent of the test material under investigation.

Each of the 11 sodium test solutions, the 8 creatinine test solutions, and the 10 or 9 thyrotropin test solutions were analysed 10 times "within run", and the average values and standard deviations calculated from the results.

The 10 identical specimens in each 10 fold "within run" determination were analysed directly one after the other, so that the results were not affected by carry-over.

2.4.1. *Average value ($\bar{x}$) and standard deviation (s)* were calculated in the conventional way:

$$\bar{x} = (1/n) \Sigma x_i; \qquad s = [\Sigma (x_i - \bar{x})^2/(n-1)]^{0.5}$$

The relative standard deviation, i.e. the so-called coefficient of variation (CV) was normally calculated as percentage CV: $CV\% = 100\ (s/\bar{x})$.

2.4.2 *Regression lines and correlation coefficients* were estimated by principal component analysis according to *Feldmann* et al. (18) or by linear regression analysis. A computer program was available for construction of the regression line of *Passing & Bablok* (19).

2.4.3 *Bias* was determined from: Bias $= \bar{x} - \mu$. The relative bias or coefficient of bias (CB%), which is analogous to CV%, was determined from: $CB\% = 100\ (\bar{x} - \mu)/\mu$.

2.5 Graphical procedure for the evaluation of precision and accuracy, and their dependency on analyte concentration

The following graphs were produced by entering the values of $\mu$, $\bar{x}$ and s in a specially written program for a PC (Hewlett Packard 87XM).

2.5.1 $\bar{x}$ vs s.

2.5.1.1 The "*Sadler* curve" (corresponding to fig. 1a).

2.5.1.2 Transformation of 2.5.1.1 to $\bar{x}$ vs CV% (corresponding to fig. 1b).

2.5.2 $\mu$ vs $(\bar{x} - \mu)$.

2.5.2.1 The appropriate regression line according to l.c. (19).

2.5.2.2 The required interval limits; for this purpose the "interval factor" for s must be calculated externally and entered separately each time, e.g. for the bias profile: $t_\alpha/\sqrt{n}$ for the deviation profile: $t_\alpha\sqrt{1 + 1/n}$ (corresponding to fig. 1c).

2.5.2.3 Transformation of 2.5.2.2 to: $\mu$ vs $(\bar{x} - \mu)/\mu \times 100$, with the accompanying transformed intervals (corresponding to fig. 1d).

These 4 presentations were chosen in a dialogue with the computer, in which the program asks whether the abscissa should be linear or logarithmic, which concentration range should be covered, which scale or scale interval should be shown on the ordinate, etc.

According to choice, the graph can be presented with or without a grid, and many different symbols are available for representation of the points.

The plot of 2.5.1.2 represents the "precision profile", while 2.5.2.3 represents the "bias profile" or "deviation profile". Different lines (continuous, dotted, etc.) can be selected for the plot, and the accompanying invervals can be hatched. The graph is produced on a VDU or a plotter, so that multicoloured graphics are also possible.

[4]) The authors are very grateful to Prof. *J. J. Staub* and Dr. *H. Engler* (Endocrinology Division, Centre for Internal Medicine of the University of Basel) for the analyses, and to Dr. *L. Bors*, Director of Behring Hoechst (Switzerland) for providing the reagents.

Tab. 2. Data base from the evaluation of the flame photometer and the ISE system for the determination of sodium concentration/activity.

| μ (mmol/l) | Flame photometer | | | | ISE | | | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ (mmol/l) | s (mmol/l) | CV (%) | CB (%) | $\bar{x}$ (mmol/l) | s (mmol/l) | CV (%) | CB (%) |
| 87.0 | 90.06 | 0.259 | 0.29 | 3.52 | 89.7 | 0.483 | 0.54 | 3.10 |
| 91.6 | 94.71 | 0.152 | 0.16 | 3.39 | 94.3 | 0.483 | 0.51 | 2.95 |
| 96.7 | 99.39 | 0.208 | 0.21 | 2.78 | 99.3 | 0.483 | 0.49 | 2.69 |
| 102.3 | 104.90 | 0.205 | 0.19 | 2.54 | 105.0 | 0.471 | 0.45 | 2.64 |
| 108.7 | 111.88 | 0.253 | 0.23 | 2.60 | 111.1 | 0.316 | 0.28 | 2.21 |
| 116.0 | 117.74 | 0.331 | 0.28 | 1.50 | 118.2 | 0.632 | 0.53 | 1.90 |
| 124.3 | 125.42 | 0.244 | 0.19 | 0.90 | 126.6 | 0.516 | 0.41 | 1.85 |
| 133.8 | 134.40 | 0.485 | 0.36 | 0.45 | 136.1 | 0.738 | 0.54 | 1.94 |
| 145.0 | 144.48 | 0.193 | 0.14 | −0.36 | 147.3 | 0.483 | 0.33 | 1.59 |
| 158.2 | 156.50 | 0.440 | 0.28 | −1.07 | 159.3 | 0.527 | 0.33 | 0.70 |
| 174.0 | 169.66 | 0.381 | 0.22 | −2.49 | 175.3 | 0.483 | 0.28 | 0.74 |

## 3. Results

### 3.1 Determination of sodium concentrations and activities

Comparison of the results of the two analytical systems is restricted by the fact that sodium concentrations differ from $Na^+$ activities, even in pure aqueous solutions. Moreover, on no account can the results for sodium analysis in aqueous solution be used to interpret analyses in biological materials (plasma, serum, urine, etc.). Table 2 shows the theoretical values, average values and standard deviations, as well as the calculated CV% and CB%.

3.1.1 *The within-run precision*, expressed as CV%, is far less than 0.5% for flame photometry, and less than 1% for the ISE system. In flame photometry, the symmetrical variations, CB%, decrease from +3.5% to −2.5% with increasing concentration; in the ISE system, this decrease is less pronounced (from +3.1% to +0.74%).

3.1.2 *The compensatory regression line* for the plot of the ISE values against the theoretical values is shown in figure 2, together with the correlation coefficient. It can be seen that all the points lie exactly on the compensatory regression line, which for the ISE system shows a slope of 0.98 and an intercept of +4.1 mmol/l. For flame photometry, the slope is 0.99, with an intercept of 2.1 mmol/l. In both systems the correlation coefficient for the theoretical values is greater than 0.999. The regression equations and correlation coefficients of the four possible combinations are given in table 3.

3.1.3 *The graph of standard deviations against average values* is shown in figure 3a. *Sadler* curves were constructed from each set of points. As the average values increase over the measurement range, their accompanying standard deviations also increase. This increase is about 0.2 mmol/l for flame photometry, but much less for the ISE system.

In contrast, the precision profile (fig. 3b) for flame photometry is almost parallel to the concentration axis over the whole measurement range, whereas the precision profile of the ISE system describes a curve which falls from 0.55% to 0.3% with increasing $Na^+$ activity.

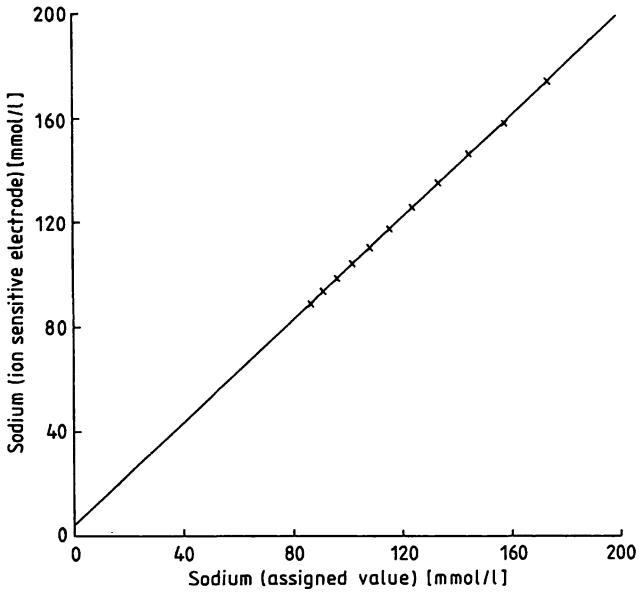


Fig. 2. Regression line (principal component analysis): assigned values (μ) vs mean values of sodium, obtained by ion sensitive electrode measurements (y).
$y = 0.984\ \mu + 4.13$, $r > 0.999$.

Tab. 3. Regression equations from the data of table 3. FlaPho = flame photometry

| Regression functions y = a x + b | | | Correlation coefficient r |
|---|---|---|---|
| y | ax | b | |
| FlaPho | = 0.987 μ | +2.13 | 0.999 |
| ISE | = 0.984 μ | +4.13 | 0.999 |
| ISE | = 1.070 FlaPho | −7.27 | 0.999 |
| FlaPho | = 0.935 ISE | +6.80 | 0.999 |

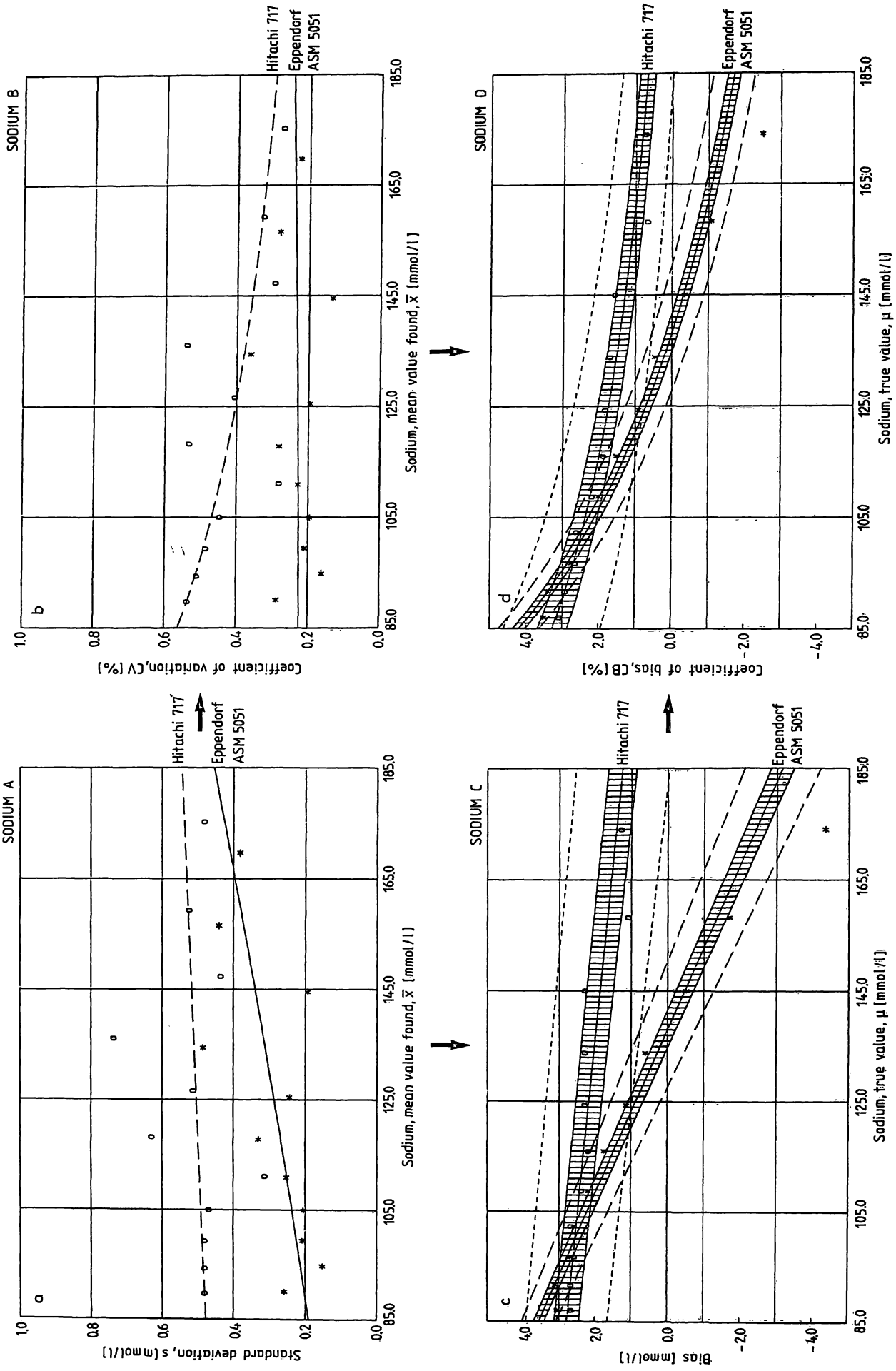


Fig. 3. Graphs of the results from the sodium determinations by flame emission spectrometry (x − x, Eppendorf ASM 5051) and ion sensitive electrode (o—o, Hitachi 717).
a) Sodium A, mean value found, x̄ vs standard deviation, s;
b) Sodium B, mean value found, x̄ vs relative standard deviation (coefficient of variation), CV;
c) Sodium C, true value, μ vs bias with intervals according to equations 3 and 4;
d) Sodium D, true value, μ vs relative bias (coefficient of bias), CB.
The bias profile (from Eq. 3) is hatched, and the tolerance limits (from Eq. 4) are shown as dashed lines.

Tab. 4. Data base for the evaluation of methods for the determination of creatinine.

| μ (μmol/l) | Hitachi 705 | | | | Manual | | | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ (μmol/l) | s (μmol/l) | CV (%) | CB (%) | $\bar{x}$ (μmol/l) | s (μmol/l) | CV (%) | CB (%) |
| 40.66 | 54.5 | 1.51 | 2.77 | 34.04 | 47.86 | 4.24 | 8.86 | 17.70 |
| 54.59 | 67.3 | 1.34 | 1.99 | 23.28 | 58.57 | 3.43 | 5.86 | 7.29 |
| 68.18 | 81.6 | 1.58 | 1.93 | 19.68 | 68.67 | 1.56 | 2.27 | 0.72 |
| 81.45 | 93.6 | 2.07 | 2.21 | 14.92 | 80.78 | 2.08 | 2.57 | −0.82 |
| 94.39 | 107.4 | 1.51 | 1.41 | 13.78 | 89.24 | 3.37 | 3.77 | −5.46 |
| 107.04 | 118.0 | 1.15 | 1.97 | 10.24 | 105.00 | 3.14 | 2.99 | −1.90 |
| 119.39 | 129.2 | 1.72 | 1.33 | 8.22 | 118.43 | 3.52 | 2.97 | −0.80 |
| 131.45 | 140.3 | 1.54 | 1.54 | 6.73 | 129.92 | 3.89 | 2.99 | −1.16 |

3.1.4 *The plot* of the differences ($\bar{x}_i - \mu_i$) (i.e. *the bias*) against the theoretical values (fig. 3c) shows a continuous linear decrease from about +3.0 mmol/l to −3.0 mmol/l for flame photometry. In contrast, the bias of the ISE system decreases only slightly, i.e. from 2.7 to 1.3 mmol/l. However, the intervals (calculated according to Eq. 3 and Eq. 4) for flame photometry are markedly narrower than those for the ISE system.

The relative deviations of accuracy are shown in figure 3d. Both analytical methods produce gently decreasing convex curves, which are more pronounced for the flame photometer than for the ISE system. Below 145 mmol/l, the two profiles increasingly overlap, whereas above 145 mmol/l they become increasingly differentiated.

## 3.2 Determination of creatinine

Table 4 shows the theoretical values, together with the average values, standard deviations and CV% and CB% values produced by each of the tested analytical systems.

3.2.1 It can be seen that the lowest *within-run precision*, expressed as CV%, is obtained with the Hitachi 705. With both systems, the variation coefficient decreases with increasing creatinine concentration. The systematic variations, CB%, also decrease with increasing creatinine concentration. For both systems, the percentage systematic variations are much greater than the percentage imprecisions.

3.2.2 *The regression line and correlation coefficient* of the manual method are shown in figure 4. The analytical values are narrowly scattered around a regression line, which has a slope of 0.91 and an intercept of +7.8 μmol/l. Table 5 shows the regression equations and correlation coefficients, the latter being greater than 0.99 for all three systems.

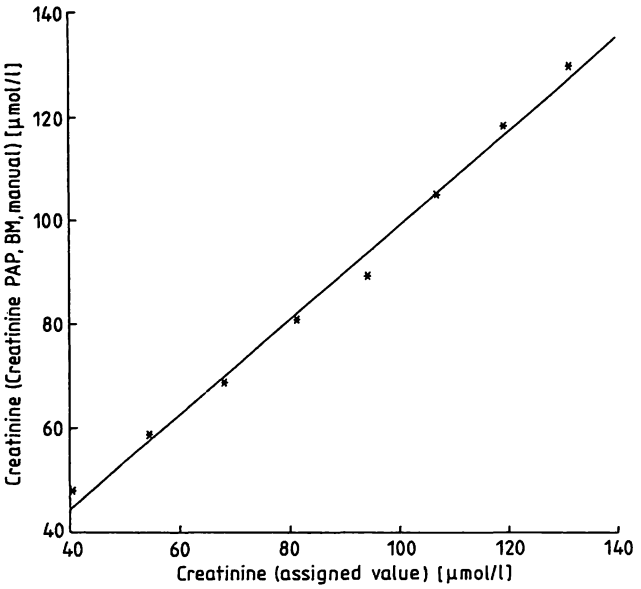


Fig. 4. Regression line (principal component analysis): assigned values (μ) vs mean values of creatinine, obtained by Creatinine PAP, Boehringer Mannheim, manual method (y).
$y = 0.912\ \mu + 7.85$, $r = 0.996$.

Tab. 5. Regression equations from the data of table 4. Hi = Hitachi

| Regression function: y = a x + b | | | Correlation coefficient r |
|---|---|---|---|
| y | ax | b | |
| Manual | = 0.912 μ | + 7.85 | 0.996 |
| Hi 705 | = 0.949 μ | +16.26 | 0.999 |
| Hi 705 | = 1.035 man | + 8.63 | 0.994 |
| Manual | = 0.955 Hi | − 7.20 | 0.994 |

3.2.3 Figure 5a shows a plot of *standard deviation against the average values*. In both methods, the associated *Sadler* curves increase with increasing creatinine concentration. It can be seen that the imprecision of the Hitachi system is lower than that of the manual procedure.

If these curves are transformed to relative imprecisions (fig. 5b), it becomes very clear how the relative

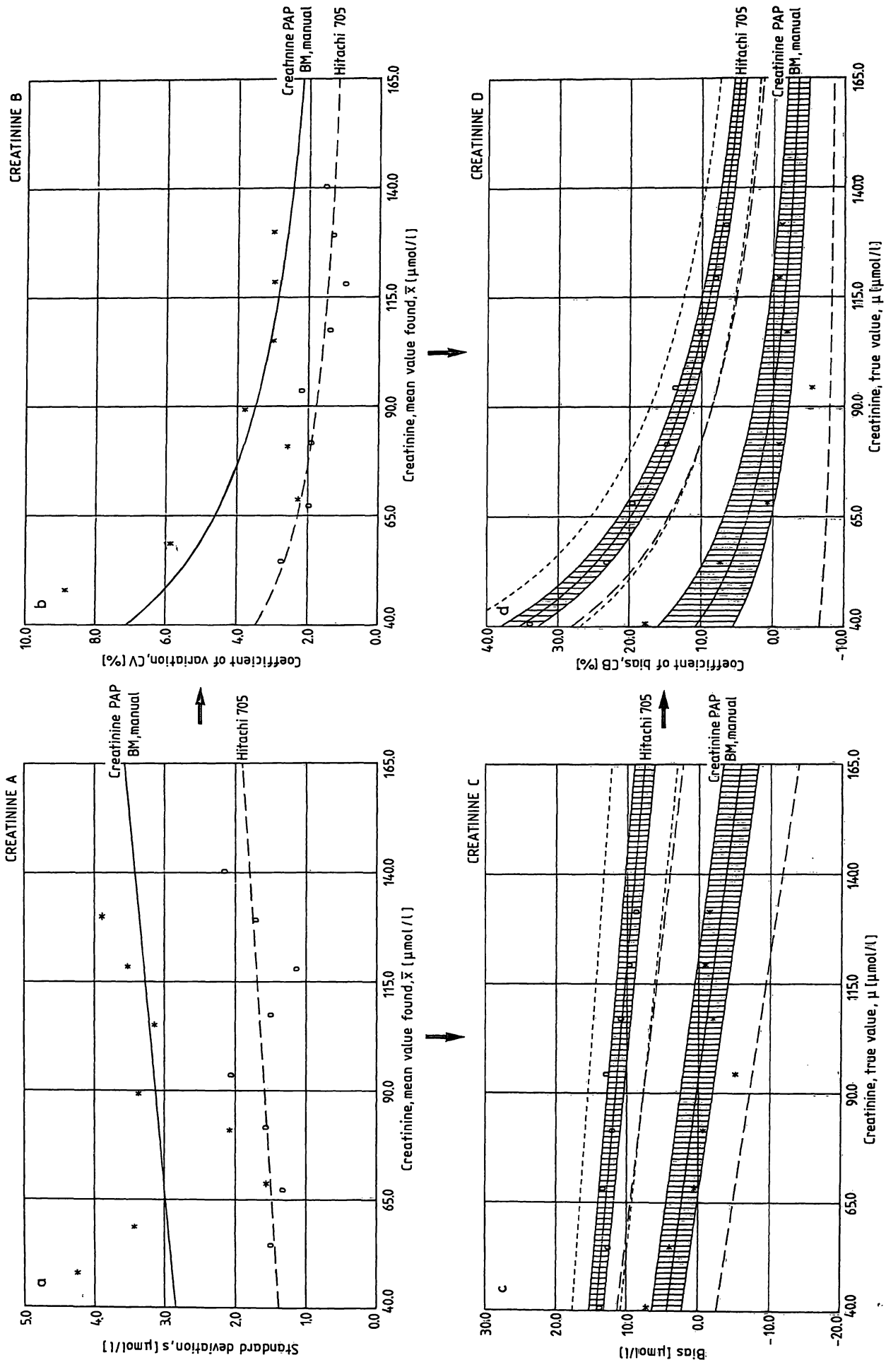


Fig. 5. Graphs of the results from the creatinine determinations by Creatinine PAP, Boehringer Mannheim, manual method ($x-\bar{x}$) and the kinetic *Jaffé* method (o—o, Hitachi 705).
a) Creatinine A, mean value found, $\bar{x}$ vs standard deviation, s;
b) Creatinine B, mean value found, $\bar{x}$ vs relative standard deviation (coefficient of variation), CV;
c) Creatinine C, true value, µ vs bias with intervals according to equations 3 and 4;
d) Creatinine D, true value, µ vs relative bias (coefficient of bias), CB.
The bias profile (from Eq. 3) is hatched, and the tolerance limits (from Eq. 4) are shown as dashed lines.

imprecision of each system decreases with increasing creatinine concentration. The percentage variation coefficient is generally less than 2.5% for the Hitachi, and generally less than 5% for the manual procedure.

3.2.4 Figure 5c shows the *absolute difference between average and theoretical values* plotted against the theoretical values. In both methods, the regression line of these differences tends to fall with increasing creatinine concentrations. This graphical representation shows that the manual method is more accurate than the Hitachi system in the investigated concentration range. Conversely, the points in the plot for the Hitachi system are less widely scattered. The lower limit of the intervals for the Hitachi system is practically identical with the upper limit fo the intervals for the manual method.

In figure 5d, the relative differences between average and theoretical values are plotted against the theoretical values. The relative *Sadler* curves are convex, and they fall with increasing creatinine concentration. Their associated deviation limits decrease with increasing creatinine concentration. Between 90 µmol/l and 165 µmol/l, the deviation limits of the manual method are less than ± 10%. The analytical values of the Hitachi system are systematically higher than the nominal values determined by HPLC. The results of the manual method, however, show a close correspondence with the target values, notwithstanding the fact that they also show a relative increase with decreasing creatinine concentration.

## 3.3 Determination of thyrotropin concentration

Table 6 shows the the theoretical values, average values and standard deviations of the two tested immunoassays.

3.3.1 Over the tested range, the *standard deviation* of both methods increases with increasing thyrotropin concentration. The lowest standard deviation was always found with the lowest thyrotropin concentration, and the largest standard deviation with the highest thyrotropin concentration.

3.3.2 The results of the radioimmunoassay are plotted against those of the reference method in figure 6, which also shows the correlation coefficient. Only three points are shown; the remaining points lie below 1 mU/l, and they are therefore invisible. Table 7 shows the regression functions for the two tested assays. Although the gradient of the RIA is about 1, and that of the LIA is only 0.6, both methods show very high correlation coefficients.

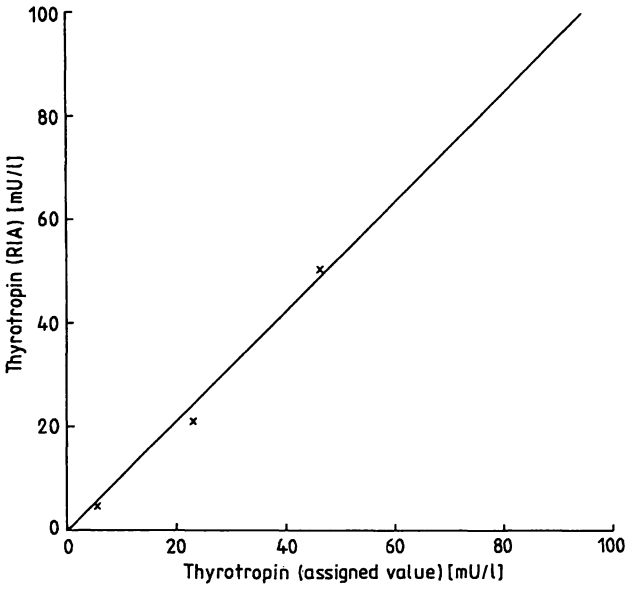


Fig. 6. Regression line (principal component analysis): assigned values (µ) vs mean values of thyrotropin obtained by radioimmunoassay (y).
$y = 1.062\ \mu - 0.265$, $r = 0.997$.

Tab. 6. Data base for the evaluation of methods for the determination of thyrotropin.

| µ (mU/l) | RIA | | | | LIA | | | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ (mU/l) | s (mU/l) | CV (%) | CB (%) | $\bar{x}$ (mU/l) | s (mU/l) | CV (%) | CB (%) |
| 0.016 | 0.005 | 0.0108 | 216 | −68.75 | 0.068 | 0.01 | 14.71 | 325 |
| 0.0341 | 0.016 | 0.018 | 112.5 | −52.79 | 0.085 | 0.013 | 15.29 | 149.26 |
| 0.0523 | 0.067 | 0.020 | 29.85 | 28.11 | 0.079 | 0.013 | 16.46 | 51.05 |
| 0.0886 | 0.088 | 0.010 | 11.36 | − 0.68 | 0.085 | 0.020 | 23.52 | − 4.06 |
| 0.1611 | 0.142 | 0.021 | 14.79 | −11.85 | 0.138 | 0.037 | 26.81 | −14.34 |
| 0.3063 | 0.244 | 0.021 | 8.61 | −20.34 | 0.178 | 0.036 | 20.22 | −41.89 |
| 0.5965 | 0.445 | 0.021 | 4.72 | −25.39 | 0.348 | 0.058 | 16.67 | −41.66 |
| 5.8215 | 4.927 | 0.055 | 1.12 | −15.37 | 3.591 | 0.238 | 6.63 | −38.31 |
| 23.238 | 21.27 | 0.402 | 1.89 | − 0.08 | ./. | ./. | ./. | ./. |
| 46.46 | 50.632 | 1.155 | 2.28 | 8.98 | 30.791 | 2.059 | 6.69 | −33.73 |

Tab. 7. Regression equations from the the data of table 6.

| Regression function: y = a x + b | | | Correlation coefficient r |
|---|---|---|---|
| y | ax | b | |
| RIA | = 1.0621 μ | −0.2646 | 0.997 |
| LIA | = 0.6623 μ | −0.1043 | 0.999 |
| LIA | = 0.6068 RIA | +0.1157 | 0.999 |
| RIA | = 1.6476 LIA | −0.1896 | 0.999 |

3.3.3 The *Sadler* curves for the plots of *standard deviations* against the average values are shown in figure 7a. Up to an average value of 1 mU/l, these curves run strictly parallel to the abscissa, and are more or less identical. Above this value, however, they rise rapidly, and at 10 mU/l they are distinctly separated. Both curves show a good fit to the points, and their apparent massive rise with increasing thyrotropin concentration is due to the logarithmic scale of the abscissa.

In contrast, the precision profiles, as expected, decrease with increasing thyrotropin concentration (fig. 7b). After passing through a minimum, they both increase again slightly (the LIA rather more than the RIA) above 10 mU/l. At 0.1 mU/l, the precision profile of the RIA is still less than 20%, whereas the precision profile fo the RIA is greater than 30%. At concentrations greater than 1.0 mU/l, the precision profiles of both assays lie below 10%.

3.3.4 The *absolute difference* between reference values and experimental results are plotted against theoretical values in figure 7c. Up to 1 mU/l, the plots for the two assays are more or less identical, and they show extremely close agreement with the reference method. Above 1 mU/l, the plots slope downwards, as both methods increasingly diverge from the reference method.

The relative deviations of accuracy are shown in figure 7d. Below 1 mU/l, the imprecision of both assays increases exponentially. Below 0.1 mU/l, both assays produce useless values. On the other hand, both assays show their narrowest deviation ranges at about 5 mU/l, and these increase markedly again at higher concentrations. Between 1 and 100 mU/l, the RIA produces values that are systematically slightly less than those of the reference method, whereas the values from the LIA are about 40% lower than the reference values.

## 4. Discussion

For the evaluation of analytical methods, *Büttner* et al. (16) put forward the IFCC recommendations. These include procedures for the determination of analytical precision and accuracy. It is stressed that the imprecision of the lower, intermediate and upper parts of the analytical range should be considered separately, and that correlation coefficients and regression equations are of little value for comparing the inaccuracy of two methods. Instead, it is recommended that two methods be compared by plotting the differences between test and reference method against the average values.

Similar approaches are formulated in the ECCLS recommendation for the "Evaluation of Analyses in Clinical Chemistry" (20) and in the NCCLS recommendations for "Comparison of quantitative clinical laboratory methods" (21).

In practically all contemporary descriptions of methods, or reports of method comparisons, data are presented on imprecision (intra-assay, between-day, etc.). These are based on measurements of two or three commercial control materials, whose concentrations usually do not lie in the upper or lower limits of detection of the method under investigation. Accuracy is also tested with commercial control materials, and the results are compared with those claimed by the suppliers. When methods are compared, the concentration of an analyte is determined with the test and reference method on a number of patient sera; the results from the respective methods are plotted against each other, a regression line is drawn, and the relevant correlation coefficient is calculated. If, as a first approximation, the points lie on the regression line, the correlation coefficient is high and the imprecision is low; the tested method is then declared to be reliable, and to show a close correlation with the reference method.

The lower and upper measurement limits are usually not determined, and no distinction is drawn between the lower detection limit and the lower limit of the analytical range. The dependency of precision and accuracy on concentration is nearly always overlooked, despite the fact that even standard text books state: "the standard deviation may vary with the level of the substance being determined (24)", and: "the random error distributions are seen to vary with the concentration of the test material ... this is a common occurrence in clinical chemical assays and is virtually the rule with radioimmunoassays" (25).

The theoretical basis of this approach is questioned, not only by the IFCC recommendations, but also by the work of *Bland* and *Altmann* (22, 23). In practical terms, the poor correlation of results from different laboratories shows that current procedures for describing and comparing methods are often inadequate.

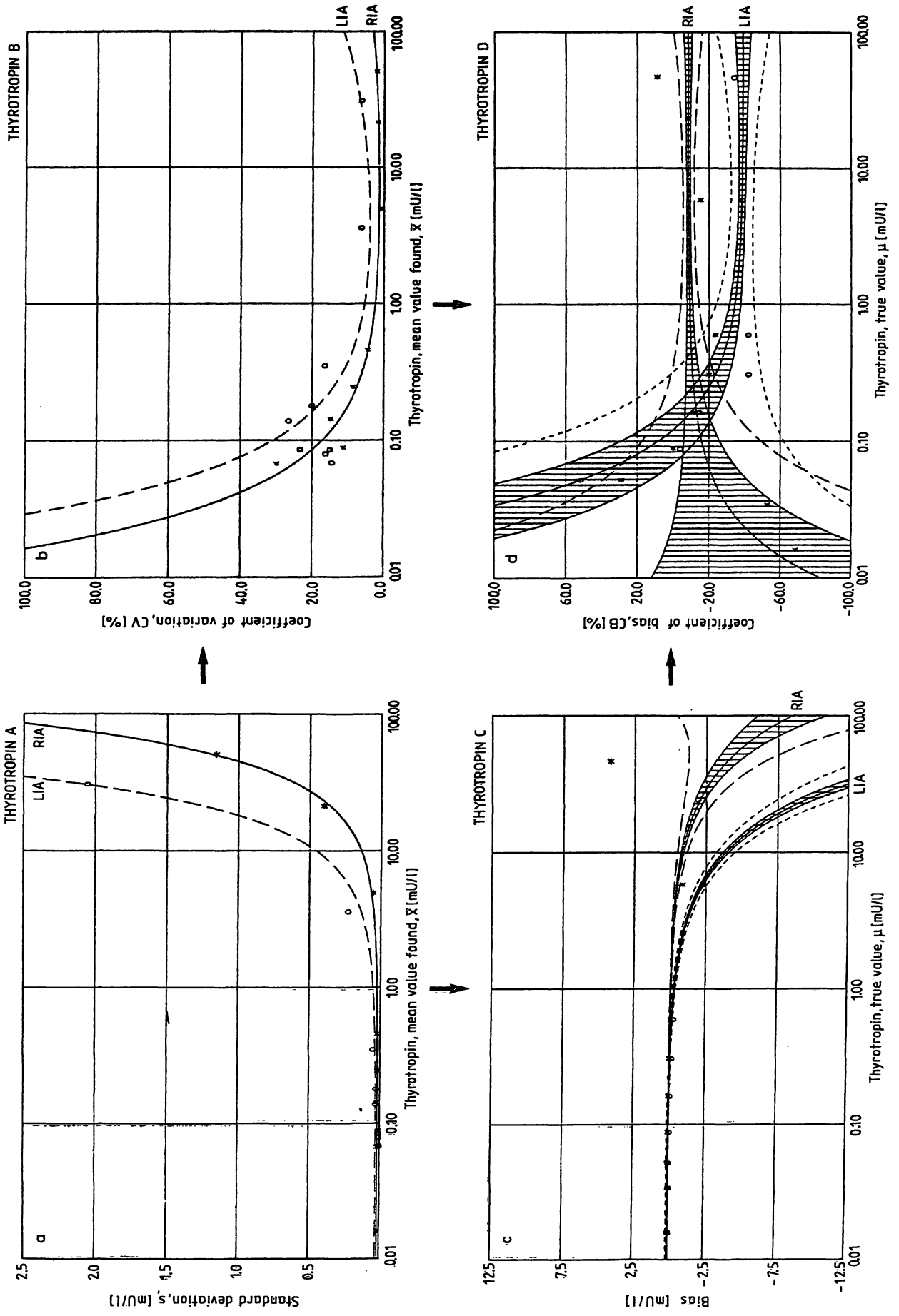


Fig. 7. Graphs of the results from the thyrotropin determinations by radioimmunoassay (x—x, RIA) and luminescence immunoassay (o—o, LIA).
a) Thyrotropin A, mean value found, $\bar{x}$ vs standard deviations, s;
b) Thyrotropin B, mean value found, $\bar{x}$ vs relative standard deviation (coefficient of variation), CV;
c) Thyrotropin C, true value, μ vs bias with intervals according to equations 3 and 4;
d) Thyrotropin D, true value, μ vs relative bias (coefficient of bias), CB.
The bias profile (from Eq. 3) is hatched, and the tolerance limits (from Eq. 4) are shown as dashed lines.

The procedure described in the present work represents an extension of the IFCC recommendations. The most important characteristics of the performance of a clinical chemical method are displayed by drawing a graph of accuracy and precision over the whole or part of the analytical range.

A combination of bias and standard deviation, as recently described by *Louderback* et al. (26), was intentionally avoided, because, as stated by *Currie* (25), these two parameters "must be viewed as two component vectors".

With the aid of the deviation profile, the lower and upper limits of the measurement range of a clinical chemical method can be clearly established by defining the limits of deviation. The limits of the measurement range are then defined by those points at which the limits of deviation are exceeded. As pointed out frequently by *Stamm* (27, 28), these limits must make sense in terms of the clinical purposes of the assay. Neither "the dynamic range of system response" (21), nor the customary estimation of (deviations from) linearity, are appropriate criteria for establishment of the measurement range.

Without doubt, the described experimental procedure is demanding and occasionally difficult to realize. This is clearly shown by the evaluation of the methods for the determination of sodium concentration and activity. Patient sera with very low or very high sodium concentrations, which would be suitable for preparation of the dilution series, are seldom encountered. We therefore resorted to an aqueous calibrator. The results therefore do not represent those obtained with biological matrices, and despite the fact that they show splendid reproducibility, they are only valid as a model.

The comparative studies on the determination of creatinine and thyrotropin are nearer to reality. Even here, there should be an awareness of the fact that changes occur in the matrix whenever two sera are mixed to produce a new test concentration of analyte. It can also be argued that only the intra-assay imprecisions are used. Determination of day-to-day imprecision is, however, very time-consuming, so that multicentre evaluations are probably more appropriate. Each test laboratory would then prepare its own pool sera and its own dilution series.

In the present investigation, all regression analyses revealed a linear relationship between the test and reference method, and all correlation coefficients were greater than 0.99. These correlation coefficients are, however, artificially high, because the x/y values are not normally distributed over the whole scale.

On the other hand, the graphical representation of the bias and deviation profiles gave a different picture. Thus, the results from the flame photometer indeed show high precision, but they become very inaccurate in the upper and lower part; the ISE results are less precise, but have a smaller bias.

The mechanized creatinine determination has a very low imprecision, but it deviates considerably from the true values, especially in the lower range. Better results are obtained with the manual, enzymic method. Similar findings have been published by *Guder* et al. (29), who presented intra-assay- and between-day-imprecision profiles for this method, based on the results of a multicentre evaluation. The plot of their intra-assay precision profile is very similar to our figure 5b.

Analytical values produced by the tested method for thyrotropin sometimes differed considerably from those of the reference method. More significantly, however, the imprecision and inaccuracy increase below 1 mU/l, until the results become useless for diagnostic purposes below 0.1 mU/l. This result is confirmed by recent results of other authors, who, however, reported only precision profiles (30, 31).

What are the implications of presenting analytical accuracy and precision in the form of bias profiles and deviation profiles?

In accordance with the ECCLS recommendations, a comparison of methods, or the investigation of the performance of a method, even with the aid of bias profiles and deviation profiles, is valid only if performed by several laboratories in a multicentre evaluation. Different matrices and different interfering factors must be investigated, and this is achieved by using different pool sera for the preparation of test series of dilutions.

The mathematical-statistical bases for the construction of bias profiles and deviation profiles are given in the appendix.

Other characteristics of clinical chemical methodology, which have not been touched upon by the present work or discussion, but nevertheless should be investigated, are analytical specificity, the influence of interfering factors, carry-over, practicability, and operational costs in terms of time and money, etc.

**Acknowledgement**

The authors are very grateful to their coworkers in the Institute, *Silvia Bosart, Bernadette Keller* and *Heinz Ryffel*, for their painstaking work in producing the experimental results; and to *Cornelia Lüthard*, for writing the manuscript.

For helpful discussions, the authors are endebted to: Dipl. Math. *W. Bablok*, Mannheim; Prof. *R. Haeckel*, Bremen; Prof. *F. Körber*, Berlin; and Prof. *D. Stamm*, München.

## References


1. Nalimov, V. V. (1963) The application of mathematical statistics to chemical analysis. Pergamon Press, London.
2. Wilson, L. A. (1970) The performance-characteristics of analytical methods II. Talanta *17*, 31–44.
3. Keller, H. (1968) Über die konzentrationsbedingte Abhängigkeit von Streuung und Richtigkeit bei klinisch-chemischen Bestimmungen. In: *Automatisierung des klinischen Laboratoriums* (Grieser, G. & Wagner, G., eds) Schattauer-Verlag Stuttgart.
4. Broughton, P. M. G. & Annan, W. (1971) Measurement of analytical precision in clinical chemistry. Clin. Chim. Acta *32*, 433–441.
5. Aronsson, T., De Verdier, C. H. & Groth, T. (1974) Factors influencing the quality of analytical methods – a system analysis, with use of computer simulation. Clin. Chem. *20*, 738–748.
6. Rodbard, D. (1974) Statistical quality control and routine data processing for radioimmunoassays and immunoradiometric assays. Clin. Chem. *20*, 1255–1270.
7. Ekins, R. P. (1976) In: *Hormone assays and their clinical application* (Loraine, J. A. & Bell, T. T., eds) Churchill Livingstrone, Edinburgh.
8. Ekins, R. P. (1983) The precision profile: its use in assay design, assessment and quality control. In: *Immunoassays for clinical chemistry*, 2nd Ed. (Hunter, W. M. & Corrie, J. E. T., eds) Churchill Livingstone, Edinburgh.
9. Wood, W. G., Waller, D. & Hantke, U. (1985) An evaluation of six solid-phase thyrotropin (TSH) kits. J. Clin. Chem. Clin. Biochem. *23*, 461–471.
10. Evans, M. C. (1988) Ten commercial kits compared for assay of thyrotropin in the normal and thyrotoxic range. Clin. Chem. *34*, 123–127.
11. Wood, W. G. (1987) Sensitive immunoassays for the determination of thyrotropin levels in serum. In: *Thyrotropin* (Leb, G., Passath, A., Eber, D. & Höfler, H., eds) W. de Gruyter Verlag Berlin.
12. Clark, P. M. S. & Price, P. C. (1986) Enzyme amplified immunoassay: A new ultrasensitive assay of thyrotropin evaluated. Clin. Chem. *32*, 88–92.
13. Sadler, W. A., Smith, M. H. & Legge, H. M. (1988) A method of direct estimation of imprecision profiles, with reference to immunoassay data. Clin. Chem. *34*, 1058–1061.
14. Mandel, J. (1978) Accuracy and precision: Evaluation and interpretation of analytical results. In: *Treatise on Analytical Chemistry*, Part 1, Vol. 1, 2nd Ed. (Kolthoff, J. M. & Elring, P. J., eds) John Wiley Sons, New York.
15. Currie, L. A. (1978) Sources of error and the approach to accuracy in analytical chemistry. In: *Treatise on Analytical Chemistry*, Part I, Vol. 1, 2nd Ed. (Kolthoff, J. M. & Elring, P. J., eds) John Wiley Sons, New York.
16. Büttner, J., Borth, R., Boutwell, H. J., Broughton, P. M. G. & Bowyer, R. C. (1980) Assessment of analytical methods for routine use. J. Clin. Chem. Clin. Biochem. *18*, 78–88.
17. Stamm, D. (1979) Recommendations for the description of a selected method in clinical chemistry. J. Clin. Chem. Clin. Biochem. *17*, 280–282.
18. Feldmann, U., Schneider, B., Klinkers, H. & Haeckel, R. (1981) A multivariate approach for the biometric comparison of analytical methods in clinical chemistry. J. Clin. Chem. Clin. Biochem. *19*, 121–137.
19. Passing, H. & Bablok, W. (1983) A new biometrical procedure for testing the equality of measurements from two different analytical methods. J. Clin. Chem. Clin. Biochem. *21*, 709–720.
20. Haeckel, R., Busch, E. W., Jennings, R. D., Kokholm, G. & Truchaud, A. (1986) Guidelines for the evaluation of analysers in clinical chemistry. ECCLS Document Vol. 3, No 2, June 1986.
21. Kennedy, J. W. (Chairholder) (1986) User comparison of quantitative clinical laboratory methods using patient samples. NCCLS publication EP9P, Villanova, Pa, 1986.
22. Bland, J. M. & Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet *i*, 307–310.
23. Altman, D. G. & Bland, J. M: (1983) Measurement in medicine: the analysis of method comparison studies. Statistician *32*, 307–317.
24. Reed, A. H. & Henry, R. J. (1974) Accuracy, precision, quality control, and miscellaneous statistics. In: *Clinical Chemistry – Principles and Technics* 2nd Ed. (Henry, R. J., Cannon, D. C. & Winkelmann, J. W., eds) Harper and Row Publ., Hagerstown, Md.
25. Strike, P. W. (1981) *Medical laboratory statistics*. Wright PSG, Bristol.
26. Louderback, A. L. & Szatkowski, P. R. (1980) The coefficient of analysis, a new figure of merit of laboratory performance. Clin. Chem. *26*, 774–775.
27. Stamm, D. (1982) A new concept for quality control of clinical laboratory investigations in the light of clinical requirements and based on reference method values. J. Clin. Chem. Clin. Biochem. *20*, 817–824.
28. Stamm, D. (1988) How the reliability of clinical laboratory analyses affects the medical assessment of results. A new model and its use. 4th Asian-Pacific Congr. Clin. Biochem. Aug. 28–Sept. 2, 1988, Hong Kong.
29. Guder, W. G., Hoffmann, G. E., Hubbuch, A., Poppe, W. A., Siedel, J. & Price, C. P. (1986) Multicentre evaluation of an enzymatic method for creatinine determination using a sensitive colour reagent. J. Clin. Chem. Clin. Biochem. *24*, 889–902.
30. Wilke, T. J. & Utley, D. J. (1988) The role of precision in determining the performance of a thyrotropin assay in diagnosing hyperthyreoidism. Clin. Chem. *34*, 1464–1467.
31. Bayer, M. F. (1987) Performance criteria for appropriate characterization of "highly sensitive" thyrotropin assays. Clin. Chem. *33*, 630–631.



Prof. Dr. Dr. H. Keller
Institut für Klinische Chemie
und Hämatologie
des Kantons St. Gallen
Frohbergstraße 3
CH-9000 St. Gallen


## Appendix

In the present work it is essential to distinguish between confidence intervals and tolerance intervals. Both terms are therefore discussed here in context, using the terminology of DIN (32, 33, 34).

Let the true concentration of the analyte in the sample be $\mu$, and let the expected value, as determined by the chosen method, be EX. The theoretical bias of the method is then EX – $\mu$, and this represents the systematic deviation. Whereas $\mu$ is known a priori,

EX is unknown and must therefore be estimated by independent measurements.

If $X_1, ..., X_n$ represent independent, normally distributed, random variables, then we write:

$$X_1, ..., X_n \sim N(EX, \sigma^2)$$

A confidence interval (C) is always aimed at a parameter of the underlying distribution, or at a quantity derived from it, e. g. at the bias $EX - \mu$. C is calculated from $X_1, ..., X_n$, and it contains the unknown quantity with a predetermined probability $1 - \alpha$:

$$P\{EX - \mu \in C | X_1, ..., X_n\} = 1 - \alpha$$

According to Eq. 3, C is given by:

$$(\bar{X} - \mu) \pm t_\alpha \cdot s \cdot \sqrt{\frac{1}{n}}$$

It is seen that the width of C decreases towards zero as n increases.

If X is an analytical value determined with the method under test, then $X - \mu$ represents the total deviation of this measurement:

$$X - \mu = (X - EX) + (EX - \mu)$$

i. e. the total deviation of the measurement consists of the sum of random deviation and the bias of the method.

A tolerance interval, T, is always an estimated range, in which the random variable, X, or a variable derived from it, e. g. $X - \mu$, varies in accordance with the distribution of X. T is also calculated from $X_1, ..., X_n$, and it contains $X - \mu$ with a predetermined expected probability of $1 - \alpha$:

$$EP\{X - \mu \in T | X_1, ..., X_n\} = 1 - \alpha$$

(Tolerance intervals with confidence probability will not be discussed here). According to Eq. 4, T is given by:

$$(\bar{X} - \mu) \pm t_\alpha \cdot s \cdot \sqrt{1 + \frac{1}{n}}$$

The width of T does not tend to zero.

Let $\Delta$ be a predetermined limit of deviation, i. e. a deviation that is still acceptable for practical purposes.

The associated measurement range then consists of all those concentrations for which the following relationship is valid:

$$|(\bar{X} - \mu) \pm t_\alpha \cdot s \cdot \sqrt{1 + \frac{1}{n}}| \leq \Delta$$

In order to clarify further the differences between C and T, the derivation of the formulae for C and T should be considered. The derivation starts with:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(EX, \frac{\sigma^2}{n}\right)$$

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Now, let t be a *Student*-distributed random variable with $(n - 1)$ degrees of freedom, and let $t_\alpha$ be its two-sided critical value for probability $1 - \alpha$; then:

$$\bar{X} - EX \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\sqrt{n} \cdot \frac{\bar{X} - EX}{s} \sim t$$

Since $\bar{X} - EX = (\bar{X} - \mu) - (EX - \mu)$, it follows that

$$P\left\{|(\bar{X} - \mu) - (EX - \mu)| \leq t_\alpha \cdot s \cdot \sqrt{\frac{1}{n}}\right\} = 1 - \alpha$$

The confidence limits for $EX - \mu$ are therefore

$$(\bar{X} - \mu) \pm t_\alpha \cdot s \cdot \sqrt{\frac{1}{n}}$$

Further:

$$\bar{X} - X \sim N\left(0, \sigma^2 (1 + \frac{1}{n})\right)$$

$$\frac{\bar{X} - X}{s \cdot \sqrt{1 + \frac{1}{n}}} \sim t$$

Since $\bar{X} - X = (\bar{X} - \mu) - (X - \mu)$ it follows that

$$P\left\{|(\bar{X}-\mu)-(X-\mu)| \leq t_\alpha \cdot s \cdot \sqrt{1+\frac{1}{n}}\right\} = 1-\alpha$$

The tolerance limits for $X - \mu$ are therefore

$$(\bar{X} - \mu) \pm t_\alpha \cdot s \cdot \sqrt{1 + \frac{1}{n}}$$

It can be seen that for $n \geq 10$, these limits show little dependence on n only. This is important, because it is obvious that the limits of an analytical range must not depend on n.

## Additional references


32. DIN 55350 Teil 12: Begriffe der Qualitätssicherung und Statistik, Merkmalsbezogene Begriffe. Entwurf April 1986.
33. DIN 58936 Teil 1: Qualitätssicherung in der Laboratoriumsmedizin, Grundbegriffe. Entwurf Dezember 1986.
34. DIN 58936 Teil 2: Qualitätssicherung in der Laboratoriumsmedizin, Begriffe zur Qualität und Anwendung von Klassierungs-, Zähl- und Meßsystemen. Entwurf Dezember 1987.